

FIA Prototype

Kamal, Chris, Liam

Components to Prototype

- Refer to the Risk Mitigation section in the Software Development Plan
 - Our most prominent concern is working with Tesseract. It is an outside source we cannot control, and all the decent search algorithms in the world cannot help if Tesseract prints out garbage we can't account for. Learning it inside and out and what we need to do to retrieve accurate text is our primary goal.
- Identify the hardware and external software your product will use
 - Tesseract OCR
 - SQL Database
 - Mobile Devices
- To this end, the prototype uses a very barebones version of a final SQL database that will eventually contain all the ingredients we need and values. This will be cross referenced during testing of test case images.

Requirements

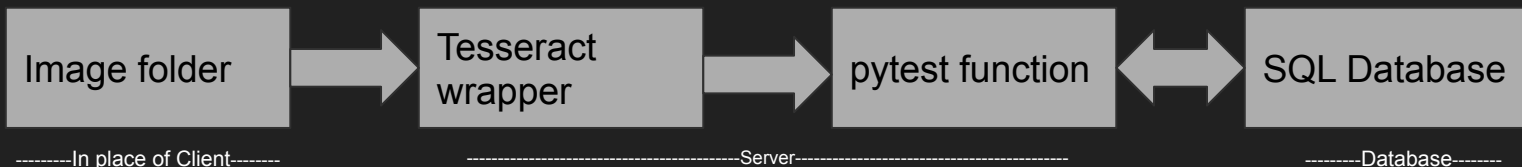
1. New requirement: SQL Database
 - a. Food ingredients and relevant dietary types pertaining to those ingredients
 - b. User preferences and blacklist
 - c. User login information
- Identify the component those requirements will be mapped to
 - SQL Database
 - Server backend

Effort Plan

- Use the schedule in Software Development Plan as a starting point
- Create detailed tasks for completing the prototype
 - Obtain test case images of various ingredient lists
 - Create a test SQL database of ingredients within those lists and whether each item is vegan or not.
 - Install Tesseract and begin implementing PIL and Tesseract for imaging and postprocessing
 - Retrieve a string from image with Tesseract, process via Python Replace() and Split() to rearrange information into useable ingredientList list[] type.
 - Connect to SQL and review each item in the ingredientList[] to be either vegan or not.
 - Print the relevant information in a digestible fashion.

Prototype Design

- Create a graphical representation



- Define interface communications; systems tend to fail at integration points
 - Tesseract retrieves image from a test case library, then returns an output of text to process.
 - Processed text is analyzed against a SQL database of vegan foods, to determine if that food is vegan or not.
- In the finished product, an HTML client will be sending images to the server. In place of this the server will retrieve test cases from a curated test image folder.

Components

- Hosting
 - Mysql server, currently on personal machine for prototyping purposes.
- Language
 - Python. Implemented pytesseract Tesseract wrapper, can interact with SQL well.
- Tools - git, third party applications, etc.
 - Git, Tesseract.
 1. Tesseract is a popular OCR (Ocular Character Recognizer) to convert Image to Text.

Test Cases 1

- Success and failure cases
 - Success, for the prototype, is successfully reading multiple images and analyzing a vegan database, to then print whether each ingredient is vegan or not.
 - If the prototype can successfully read with a 80-85% clarity, it should be accurate enough to then work on post processing first the image and then the text results for further accuracy.
 - Failure is if the image cannot be read, or if the image is processed and returns a less than 80% accurate result.

Test Cases 2

- Test 1: ingredients.py
 - Test 1 is with a perfectly clear image, with unreasonably high standards of image quality. We are testing Tesseract and PIL's latent postprocessing, and what needs to be done for each text list to retrieve understandable results.
 - Input: ingredients.py
 - Output: list of ingredients from ingredients.py, in nl delimited list, stating whether each ingredient is vegan or not.
 - Success: Yes. Ingredients.py returns

INGREDIENTS: SUGAR, ENRICHED BLEACHED FLOUR (WHEAT FLOUR, NIACIN, REDUCED IRON, THIAMIN MONONITRATE, RIBOFLAVIN, FOLIC ACID), SEMI-SWEET CHOCOLATE CHIPS (SUGAR, CHOCOLATE LIQUOR, COCOA BUTTER, SOY LECITHIN [EMULSIFIER], VANILLA), COCOA (PROCESSED WITH ALKALI), CANOLA OR SOYBEAN OIL, BITTERSWEET CHOCOLATE CHIPS (CHOCOLATE LIQUOR, SUGAR, COCOA BUTTER, MILK FAT, SOY LECITHIN [EMULSIFIER], VANILLA), MILK CHOCOLATE CHIPS (SUGAR, WHOLE MILK POWDER, CHOCOLATE LIQUOR, COCOA BUTTER, SOY LECITHIN [EMULSIFIER], VANILLA), SALT, ARTIFICIAL FLAVOR, SODIUM BICARBONATE.

SUGAR may be vegan.
ENRICHED BLEACHED FLOUR may be vegan.
SEMI-SWEET CHOCOLATE CHIPS is not vegan.
COCOA may be vegan.
CANOLA may be vegan.
SOYBEAN OIL may be vegan.
BITTERSWEET CHOCOLATE CHIPS is not vegan.
MILK CHOCOLATE CHIPS is not vegan.
SALT may be vegan.
ARTIFICIAL FLAVOR may be vegan.
SODIUM BICARBONATE may be vegan.

Test Cases 3

- Test 1: ingredients2.py
 - Test 2 is with a reasonably clear image, with larger than average content of parenthesis, and high volume of text. We are testing how well image processing can be done on dense images with a large amount of special characters.
 - Input: ingredients2.py
 - Output: list of ingredients from ingredients2.py, in nl delimited list, stating whether each ingredient is vegan or not.
 - Success: Yes. The printout is to the desired 80% accuracy rate.

INGREDIENTS: Maltodextrin, Rice Protein, Oat Flour, Vitamin and Mineral Blend [Potassium (as Potassium gluconate), choline bitartrate, Calcium (as calcium carbonate), Magnesium (as magnesium oxide), Vitamin C (as ascorbic acid), Vitamin E (as dl-alpha-tocopheryl acetate), Vitamin B3 (as niacinamide), Zinc (as zinc sulfate), Vitamin B5 (as calcium D-pantothenate), Vitamin B6 (as pyridoxine HCL), Copper (as copper gluconate), Manganese (as manganese sulfate), Vitamin B2 (as riboflavin), Vitamin A (as palmitate), Vitamin B1 (as thiamin HCL), Folic Acid , Biotin, Iodine (as potassium iodide), Chromium (as chromium chloride), Vitamin K1 (as phytonadione), Selenium (as sodium selenite), Vitamin D2 (as ergocalciferol), Molybdenum (as sodium molybdate), Vitamin B2 (as riboflavin)], Gum Acacia, Soybean Lecithin, Salt, Artificial Flavor, Sucralose.

INGREDIENTS: Canola Oil, Fish Oil

Contains: Soy

Maltodextrin is not in the database!
Rice Protein is not in the database!
Oat Flour is not in the database!
Vitamin and Mineral Blend [Potassium is not in the database!
choline bitartrate is not in the database!
Calcium is not in the database!
Magnesium is not in the database!
Vitamin C is not in the database!
Vitamin E is not in the database!
Vitamin B3 is not in the database!
Zinc is not in the database!
Vitamin B5 is not in the database!
Vitamin B6 is not in the database!
Copper is not in the database!
Manganese is not in the database!
Vitamin B2 is not in the database!
Vitamin A is not in the database!
Vitamin B1 is not in the database!
Folic Acid is not in the database!
Biotin is not in the database!
Iodine is not in the database!
Chromium is not in the database!
Vitamin k1 is not in the database!
Selenium is not in the database!
Vitamin D2 is not in the database!
Molybdenum is not in the database!
Vitamin B2 is not in the database!
Gum Acacia is not in the database!
Soybean Lecithin is not in the database!
Salt may be vegan.
Artificial Flavor may be vegan.
Sucralose Canola Oil is not in the database!
Fish Oil Contains: Soy is not in the database!

Test Cases 4

- Test 1: ingredients3.py
 - Test 3 is with a reasonably clear image, with relatively high standards of image quality. We are testing Tesseract with an image that is more average, with decent spread of special characters and lower image quality.
 - Input: ingredients3.py
 - Output: list of ingredients from ingredients3.py, in nl delimited list, stating whether each ingredient is vegan or not.
 - Success: Yes. Ingredients3.py returns a reasonably accurate list of ingredients.

INGREDIENTS: Enriched unbleached flour (wheat flour, malted barley flour, ascorbic acid [dough conditioner], niacin, reduced iron, thiamin mononitrate, riboflavin, folic acid), sugar, degermed yellow cornmeal, salt, leavening (baking soda, sodium acid pyrophosphate), soybean oil, honey powder, natural flavor.

CONTAINS: Wheat.

May contain milk, eggs, soy and tree nuts.

sodium acid —pyrophos hate soybean oil is not in the database!
reduced iron is not in the database!
honey powder is not in the database!
Enriched unbleached flour is not in the database!
ea takin soda is not in the database!
sugar may be vegan.
~— fiboflavin is not in the database!
malted barley flour is not in the database!
folic acid is not in the database!
eggs is not in the database!
soy and tree nuts is not in the database!
degermed yellow cornmeal is not in the database!
natural flavor CONTAINS: Wheat May contain milk is not in the database!
ascorbic acid is not in the database!
wheat flour is not in the database!
salt may be vegan.
dough conditioner is not in the database!
niacin is not in the database!
thiamin mononitrate is not in the database!